**APPROVAL SHEET**

This Thesis Project entitled \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

submitted by \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ in partial fulfillment of the requirements for the degree of **Bachelor of Science in Computer Science** has been examined and is recommended for acceptance for approval for ORAL EXAMINATION.

**Denise Lou B. Punzalan, MSCS**

Adviser

**PANEL OF EXAMINERS**

Approved by the Committee on Oral Examination with a grade of \_\_\_\_\_\_\_\_\_\_\_

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Erlinda S. Casela – Abarintos, DIT** | | |  |
|  | College Dean | | |  |
|  | | | | |
| **Ronnie D. Luy, MIT** | |  | **Denise Lou B. Punzalan, MSCS** | |
| Assistant Dean | |  | Research Coordinator | |

|  |  |  |
| --- | --- | --- |
|  | | |
| **Loudel M. Manaloto** |  |  |
| Program Coordinator |  | Project Adviser |

Accepted and approved in partial fulfillment of the requirements for the degree of **Bachelor of Science in Computer Science.**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date of Final Oral Defense

**ABSTRACT**

Developing a Simple Summarization Tool Using K

Nearest Neighbors (KNN) and Comparing it with

GPT-4 for Summarizing Educational Materials

By

Xyrell Dave Pamintuan

Roescen Abie S. Pajaro

Bachelor of Science in Computer Science

September 16,2024

**Ms. Denise Lou B. Punzalan**

Adviser

***Keywords:***

***Summarization Tool***

***K-Nearest Neighbors (KNN)***

***GPT-4***

***Educational Materials***

***Natural Language Processing (NLP)***

***Algorithm Design***

***Efficiency***

***Accuracy***

***Usability***

***Computational Resources***

**ACKNOWLEDGEMENT**

We would like to thank our adviser,Ms. Denise Lou B. Punzalan for her guide and assistance during this project's growth. Her knowledge and direction were crucial to our research  
  
We also sincerely appreciate the valuable insights and assessments provided by the members of the Committee on Oral Examination and the Panel of Examiners. For their invaluable time and contributions to this research, we are especially grateful to Mrs. Erlinda S. Casela – Abarintos, DIT, College Dean; Sr.Ronnie D. Luy, MIT, Assistant Dean; Ms. Denise Lou B. Punzalan, MSCS, Research Coordinator; and Sr.Loudel M. Manaloto, Program Coordinator.  
  
My teammate and I, Zenith, worked together as a team, and we would like to express our gratitude to everyone invaluable time to our project.

**EXECUTIVE SUMMARY**

The ability to quickly and accurately summarize information, including textbooks, lecture notes, and research papers, students must be able to summarise content fast and precisely. Even though sophisticated models like GPT-4 produce excellent summaries, students with limited resources may find them less useful due to their high computing requirements. The goal of this project is to provide a more user-friendly summarizing tool that makes use of the K-Nearest Neighbors (KNN) algorithm and is especially made to handle instructional content on standard devices.  
  
The main goal of this project is to develop a summary tool that may be used to simplify complicated instructional content without requiring a lot of processing resources. The study aims to assess the relative performance of the KNN-based summarization tool and GPT-4 in terms of speed, accuracy, and overall user satisfaction. Surveys are going to conducted among students to gather feedback on usability and satisfaction, while technical analysis will focus on each tool's resource consumption and efficiency.

The expected outcomes include a functional, user-friendly KNN-based summarization tool and a comparative analysis highlighting the strengths and weaknesses of both approaches. This research will provide insights into the viability of simpler algorithms like KNN as practical alternatives to resource-intensive models like GPT-4 for educational use. The findings could help guide the development of more affordable, accessible tools, enhancing students' ability to focus on key information with ease.

**TABLE OF CONTENTS**

Title Page i

Approval Sheet ii

Abstract iii

Acknowledgement iv

Dedication v

Table of Contents vi

List of Tables vii

List of Figures viii

**Chapter 1**

**THE PROBLEM AND ITS BACKGROUND**

* 1. Background of the Study
  2. Objectives of the Study
  3. Review of Related Studies
  4. Theoretical and Conceptual Framework
  5. Significance of the Study
  6. Scope and Limitation of the Study

1.7 Definition of Terms

**Chapter 2**

**RESEARCH METHODOLOGY**

2.1 Research Design

2.2 Locale of the Study *(depending on the research design and data)*

2.3 Research Instrument

2.4 Data Collection *(depending on the research design and data)*

2.5 Data Preparation *(depending on the research design and data)*

2.6 Algorithm Process Evaluation

2.7 Statistical Treatment of Data

**Chapter 3**

**PRESENTATION, ANALYSIS AND INTERPRETATION OF DATA**

3.1 Data Analysis and Presentation

3.2 Summary on Interpretation of Data

**Chapter 4**

**SUMMARY, FINDINGS, CONCLUSION AND RECOMMENDATION**

4.1 Summary of Findings

4.2 Conclusions

4.3 Recommendations

**REFERENCES**

1. Books
2. Journals/Articles/Issuances /Periodicals
3. Electronic Sources

**APPENDICES**

1. Output / Outcome of the Study
2. Letter of Request to Conduct a Study
3. Letter to the Respondents
4. Questionnaire
5. Computation of Validity Test of the Questionnaire
6. Other

**CURRICULUM VITAE**

**LIST OF TABLES**

**LIST OF FIGURES**

**Chapter 1**

**THE PROBLEM AND ITS BACKGROUND**

* 1. Background of the Study

Summarization tools are essential for students who need to efficiently process large amounts of educational material, including textbooks, lecture notes, and research papers. In recent years, artificial intelligence (AI) models like GPT-4 have demonstrated remarkable ability in generating summaries. These advanced models require computational power, making them inaccessible to many students with limited resources. This project focuses on developing a simpler, more efficient summarization tool using the K-Nearest Neighbors (KNN) algorithm. KNN is a non-parametric algorithm that classifies data points based on their proximity to other points, making it less resource intensive than large scale AI models. The aim is to create a tool that runs efficiently on standard devices, offering an alternative for students who need summarization but lack access to high powered systems.

* 1. Theoretical and Conceptual Framework

This study is grounded in the theory of natural language processing, which deals with the interaction between computers and human language. Text summarization, as a subfield of NLP, involves use large bodies of text into shorter versions while retaining key information. The conceptual framework focuses on comparing two approaches to text summarization: the complex, resource-heavy Transformer-based models like GPT-4 and simpler, more accessible algorithms like KNN. The study will assess which method performs better in summarizing educational materials with lower computational demands.

1.3 Statement of the Objectives

· To develop a summarization tool using the KNN algorithm for educational materials.

· To compare the KNN-based summarization tool with GPT-4 in terms of accuracy and computational efficiency.

· To assess student satisfaction and use for both summarization tools.

· To evaluate which tool performs better in terms of speed and resource use.

· To provide recommendations for improving summarization tools for student educational purposes.

* 1. Significance of the Study

The study's significance lies in its potential to provide students with a more accessible and efficient tool for summarizing educational content.For the student easily manage their academic workloads, a lightweight summarization tool could greatly benefit those with limited computing resources. This research could pave the way for developing affordable, easy to use educational tools that enhance learning outcomes. Additionally, the comparison between KNN and GPT-4 will offer valuable insights into how simpler algorithms can serve as practical alternatives to more sophisticated models in specific educational contexts.

* 1. Scope and Limitation of the Study

The scope of this study is limited to the development and evaluation of a KNN-based summarization tool compared to GPT-4 for summarizing educational materials. It will focus on summarizing textbooks, lecture notes, and research papers relevant to students.The study will gather feedback from a sample of students to assess ease of use and satisfaction.The study is limited by the size and diversity of the student sample and by the specific types of educational content used for testing. Additionally, the study will not explore more complex machine learning models beyond KNN and GPT-4.

1.6 Definition of Terms

**K-Nearest Neighbors (KNN):** A simple algorithm used for classification and regression, which classifies data points based on their proximity to other points.

**Summarization:** The process of condensing large bodies of text into shorter versions while retaining the key information.

**GPT-4:** A large language model developed by OpenAI, capable of generating human-like text based on given prompts.

**Natural Language Processing (NLP):** A field of AI that focuses on the interaction between computers and human languages.

**Transformer Model:** A type of deep learning model known for its effectiveness in tasks like text generation and summarization, used in advanced AI models like GPT-4.

**REFERENCES**

1. Books
2. Journals/Articles/Issuances /Periodicals
3. Electronic Sources

**APPENDICES**